Impact of training method on the robustness of the visual assessment of 18F-florbetaben PET scans - results from a phase 3 study

John Seibyl,[1] Ana M Catafau,[2] Henryk Barthel,[3] Kenji Ishii,[4] Christopher C Rowe,[5] James B Leverenz,[6] Bernardino Ghetti,[7] James W Ironside,[8] Masaki Takao, [9,10] Hiroyasu Akatsu,[11] Shigeo Murayama,[12,13] Santiago Bullich,[2] Andre Mueller,[2] Norman Koglin,[2] Walter J Schulz-Schaeffer,[14] Anja Hoffmann, [15] Marwan N Sabbagh,[16] Andrew W Stephens,[2] Osama Sabri[3]

[1]Molecular Neuroimaging LLC, New Haven, Connecticut, USA; [2]Piramal Imaging GmbH, Berlin, Germany; [3]Leipzig University, Leipzig, Germany; [4]Department of Positron Medical Center, Tokyo Metropolitan Institute of Gerontology, Tokyo, Japan; [5]Austin Health, University of Melbourne, Melbourne, Victoria, Australia; [6]VA-Puget Sound Health Care System and University of Washington, Seattle, USA; [7]Indiana University School of Medicine, Indianapolis, Indiana, USA; [8]University of Edinburgh, Edinburgh, Scotland; [9]Mihara Memorial Hospital, Isesaki, Japan; [10] Department of Neurology, Saitama International Medical Center, Saitama Medical University, Saitama, Japan; [11]Fukushimura Hospital, Toyohashi, Japan, and Departments of Community-based Medicine and Neurology, Nagoya City University Graduate School of Medical Sciences, Nagoya City, Aichi, Japan; [12]Department of Neurology and Neuropathology (the Brain Bank for Aging Research), Tokyo Metropolitan Geriatric Hospital, Tokyo, Japan; [13]Institute of Gerontology, Tokyo, Japan; [14]Georg-August University Göttingen, Göttingen, Germany; [15]Bayer Pharma AG, Berlin, Germany; [16]Alzheimer's and Memory

---

This is the author's manuscript of the article published in final edited form as:
Seibyl, J., Catafau, A. M., Barthel, H., Ishii, K., Rowe, C. C., Leverenz, J. B., … Sabri, O. (2016). Impact of Training Method on the Robustness of the Visual Assessment of 18F-Florbetaben PET Scans: Results from a Phase-3 Study. Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine, 57(6), 900–906. https://doi.org/10.2967/jnumed.115.161927

Disorders Division, Department of Neurology, Barrow Neurological Institute, Phoenix, Arizona, USA

Corresponding author

Dr John Seibyl

Molecular Neuroimaging

60 Temple Street, 8A

New Haven

Connecticut

USA 06510

Funding

The trial was funded by Bayer Pharma AG, Berlin (Germany), and Piramal Imaging S.A., Matran (Switzerland).

ABSTRACT

Training for accurate image interpretation is essential for the clinical use of β-amyloid positron emission tomography (PET) imaging, but the role of reader training and the accuracy of the algorithm for routine visual assessment of florbetaben PET scans are unclear. The aim of this study was to test the robustness of the visual assessment method for florbetaben scans, comparing efficacy readouts across different readers and training methods, and against a histopathology standard of truth (SoT).

METHODS: Analysis was based on data from an international open-label, non-randomized, multicenter Phase 3 study in patients with or without dementia (ClinicalTrials.gov: NCT01020838). Florbetaben scans were assessed visually and quantitatively, and results were compared with amyloid plaque scores. For visual assessment, either in-person training (n=3 expert readers) or an electronic training method (n=5 naïve readers) was used. Brain samples from participants who died during the study were used to determine the histopathological SoT using Bielschowsky silver staining (BSS) and immunohistochemistry for β-amyloid plaques.

RESULTS: Data were available from 82 patients who died and underwent post-mortem histopathology. Comparing visual assessment results with BSS+immunohistochemistry as SoT, median sensitivity was 98.2% for the in-person trained readers and 96.4% for the e-trained readers, while median specificity was 92.3% and 88.5%, respectively. Median accuracy was 95.1% and 91.5%, respectively. Based on BSS only as the SoT, median sensitivity was 98.1% and 96.2%, respectively, median specificity was 80.0% and 76.7%, and median accuracy was 91.5% and 86.6%. Inter-reader agreement (Fleiss' kappa) was excellent (0.89) for in-person-trained readers and very good (0.71) for e-trained readers. Median intra-reader agreement was 0.9 for both in-person and e-trained readers. Visual and quantitative assessments were

concordant in 88.9% of scans for in-person trained readers and in 87.7% of scans for e-trained readers.

CONCLUSION: Visual assessment of florbetaben images was robust in challenging scans from elderly end-of-life individuals. Sensitivity, specificity, and inter-reader agreement were high, independent of expertise and training method. Visual assessment was accurate and reliable for detection of plaques using BSS and immunohistochemistry, and well correlated with quantitative assessments.

Key words: Alzheimer's disease, florbetaben, positron-emission tomography

INTRODUCTION

Florbetaben is an ^{18}F-labeled β-amyloid tracer developed for positron emission tomography (PET). In 2014, florbetaben was approved by the US Food and Drug Administration, the European Medicines Agency and the Korean Ministry of Food and Drug Safety (MFDS) to detect or exclude the presence of neuritic β-amyloid plaques in brain (*1*,*2*). In a Phase III study, high sensitivity and specificity were demonstrated for florbetaben in the detection of β-amyloid aggregates comparing PET with post-mortem histopathology (*3*).

The current clinical method for interpretation of florbetaben β-amyloid PET scans is visual assessment (*4*). Training for accurate image interpretation is a key issue for β-amyloid PET imaging – all readers require training and several different training tools are available. The type of reader training and the accuracy for visual assessment of florbetaben PET scans are unclear. In order to test the adequacy of the training comparing visual assessment against β-amyloid histopathology, scans from end-of-life patients are required. It is known that pronounced atrophy and other brain abnormalities can compromise image interpretation in end-of-life individuals. Indeed, severe structural brain abnormalities were present in the participants in the histopathology cohort, leading to challenging scans. If acceptable images can be obtained under these conditions, improved reads can be expected in the clinical setting with less technically challenging scans. Moreover, a technically difficult dataset would be likely to expose differences between training approaches (electronic training tool vs. in-person trainer) that might inform best training practices.

The aim of the present study was to test the robustness of the visual assessment method for florbetaben scans, comparing efficacy readouts (sensitivity, specificity, and kappa values) across different readers and training methods, against a histopathology standard of truth (SoT) in

end-of-life patients. The study also compared visual and quantitative assessments of florbetaben PET, since quantification is commonly used in research and may be implemented in future clinical routine.

MATERIALS AND METHODS

Study design and population

This analysis was based on an international open-label, non-randomized, multicenter Phase 3 study (ClinicalTrials.gov: NCT01020838). The study was conducted in accordance with the Declaration of Helsinki, and approvals from regulatory authorities and ethics committees were obtained.

Participants were recruited from 15 centers (including dementia clinics with brain-bank experience, hospices, private practices, and dementia self-help groups) in Australia, Europe, Asia, and North America, and examined between February 2010 and August 2013. Eligible subjects were non-demented individuals, n=9, and patients with Alzheimer's disease (AD), n=60, dementia with Lewy bodies (DLB), n=4, or other dementias, n=9. Key exclusion criteria were cerebral large-vessel disease, brain tumors, and cardiovascular instability requiring intensive care or therapeutic intervention. All participants (or their legal representatives) provided written informed consent to undergo brain MRI, a PET scan with florbetaben, and to donate their brain for post-mortem examination. Details of the study methods have been presented previously (*3*).

Brain Image Data Acquisition

PET images were acquired 90–110 minutes after intravenous injection of 300 MBq ± 20% florbetaben (*5*) according to a standardized acquisition and image-processing protocol established during a technical visit to each center. Three-dimensional volumetric T1-weighted

brain MRI (e.g. magnetization prepared rapid gradient echo or spoiled gradient recalled sequences) were collected.

Study design

Florbetaben scans were assessed visually and quantitatively, and results compared with β-amyloid presence/absence in pathology. For visual assessment, either in-person training (n=3 expert readers) or an electronic training (e-training) method (n=5 naïve readers) was used. Composite standardized uptake value ratios (SUVRs) were determined (*6*), and receiver operating characteristic curve analysis used to ascertain the optimal threshold for the sensitivity/specificity calculations. The composite SUVR providing the highest sum of sensitivity and specificity was selected as cut-off value.

Visual assessment method

The in-person expert training and electronic training modules were identical in approach and content (*6,7*). The training emphasized normal white matter anatomy using structural MRI and coregistered florbetaben PET images to appreciate white matter - gray matter boundaries since a positive scan demonstrates extension of radiotracer uptake beyond the cortical white matter to adjacent gray matter in key brain regions. Specifically, readers used a regional cortical tracer uptake scoring system (RCTU) (1 = no tracer uptake, 2 = moderate tracer uptake, 3 = pronounced tracer uptake) in four brain areas: lateral temporal cortex, frontal cortex, posterior cingulate cortex/precuneus, parietal cortex (see figure 1 for details). The resulting scores condense into a binary interpretation (score 1 = negative; score 2 or 3 = positive). A RCTU score of 1 in each brain region led to a brain amyloid plaque load (BAPL) score of 1, a RCTU score of 2 in any brain region and no score 3 led to a BAPL score of 2. A RCTU score of 3 in any of the 4 brain regions led to a BAPL of 3.

No access to other scan orientations (i.e. coronal, sagittal), and no reorientation or structural information from computed tomography (CT) or magnetic resonance imaging (MRI) was available to readers. Readers viewed scans in gray scale only. There were some minor differences between the training, as the trainees could not ask questions of an expert reader during the review with the electronic training tool.

All images were assessed by eight readers: three in-person trained experts and five naïve readers trained using an e-training tool. All readers were nuclear medicine physicians. The ”expert reader” was defined as having direct experience with PET amyloid scans. The naïve reader had no experience with the visual assessment method and was not involved in any pivotal study using an 18F-labeled amyloid imaging agent.

Quantitative assessment method

Brain PET image quantification was performed using a standardized volume of interest template applied to the spatially normalized grey matter PET image based on a gray/white/cerebrospinal fluid segmentation of the participant's T1-weighted volumetric MRI (*6*). A region of interest template (*6*) sampled the lateral temporal, frontal, anterior and posterior cingulate gyrus/precuneus, and parietal lobes which were averaged to determine a composite SUVRs calculated using the cerebellar cortex as the reference tissue. 81 of the 82 brains were evaluated in this fashion; in one scan the segmentation process failed owing to poor technical quality of the MRI.

Pathology standard of truth

Brain samples from 82 participants who died during the study were used to determine the histopathological SoT. Six brain regions were examined with both Bielschowsky silver staining (BSS) and immunohistochemistry for the β-amyloid protein: middle frontal gyrus, occipital

cortex, hippocampus/parahippocampal gyrus, anterior cingulate cortex, posterior cingulate cortex/precuneus, and cerebellar cortex.

The presence of amyloid plaques was assessed by a blinded histopathology consensus panel of three expert neuropathologists using two different methods: Bielschowsky silver staining (BSS) and immunohistochemistry for β-amyloid. For the analysis presented here only neuritic plaques and cored plaques were considered. Neuritic plaque density (as detected by BSS) was assessed according to the Consortium for Establishing a Registry for Alzheimer Disease (CERAD) criteria (*8*), providing a semi-quantitative score with the categories "absent", "sparse", "moderate", or "frequent". The same semi-quantitative categories were used to score the number of cored plaques detected by β-amyloid immunohistochemistry. β-amyloid was regarded as present in a given brain region when sufficient neuritic or cored plaques were present to achieve a score of "moderate" or "frequent".

Importantly, BSS is not specific for β-amyloid deposits and also has some technical limitations that may hinder the identification of some neuritic plaques in AD (*9*). The combination of BSS and immunohistochemistry for β-amyloid is recommended in current neuropathologic guidelines for assessment of AD pathology (*10,11*). Therefore, both BSS and BSS+immunohistochemistry data were used for further analyses.

Statistical analysis

Sensitivity, specificity and accuracy were evaluated by comparing visual assessments with the histopathological SoT. Independent Mann-Whitney tests were performed to compare the reported parameters between training groups. Reliability of visual assessment was evaluated by Cohens' and Fleiss' kappa. To investigate the intra-reader agreement, a random subsample of images (22 for the in-person trained group and 20 for the e-trained) was re-read by all readers.

The association between quantitative and visual assessments was evaluated with a Chi squared test.

RESULTS

Study population and post-mortem β-amyloid histopathology

In total, 205 end-of-life individuals underwent florbetaben PET imaging (52% male; mean age, 76.9 ± 11 [range, 48–98] years). As of August 2013, 82 participants had died and undergone autopsy and post-mortem histopathology (clinical diagnoses: AD, n=60; DLB, n=4; other dementias, n=9; non-demented, n=9). Comparison of clinical diagnosis of AD and β-amyloid pathology (BSS+immunohistochemistry as SoT) revealed a discordance in some patients with AD, with 13 of the 60 patients with a clinical diagnosis of probable AD were found to be β-amyloid negative. Of the four subjects with DLB as clinical diagnosis, two were β-amyloid positive. Seven of the nine subjects clinically classified as other dementia were β-amyloid negative. 4 of the 9 non-demented healthy elderly subjects were β-amyloid positive. In total, of the 82 brains examined by histopathology, 56 were β-amyloid positive and 26 were β-amyloid negative with BSS+immunohistochemistry as SoT. Using BSS alone, 52 brains were β-amyloid positive and 30 β-amyloid negative.

Sensitivity and specificity of the visual assessment

Of the 82 PET scans assessed in this study, 81 were classified equally by both training groups. With BSS+immunohistochemistry as SoT, a median (range) sensitivity of 98.2% (94.6–98.2%) was obtained for the three in-person trained readers and 96.4% (91.1–100%) for the five e-trained readers. With BSS alone, median (range) sensitivity was 98.1% (96.2–98.1) and 96.2% (90.4–100%), respectively (Table 1). With BSS+immunohistochemistry, median (range) specificity was 92.3% (88.5–92.3%) for the in-person trained readers and 88.5% (53.9–92.3%)

for the e-trained readers, whereas BSS alone led to 80.0% (76.7–83.3%) and 76.7% (46.7–80%), respectively (Table 1). Median (range) accuracy with BSS+immunohistochemistry was 95.1% (93.9–96.3%) for in-person trained readers and 91.5% (84.1–93.9%) for e-trained readers. BSS alone as SoT had a median accuracy of 91.5% (90.2–91.5%) for in-person trained readers and 86.6% (79.3–89%) for e-trained readers (Table 1). Independent Mann-Whitney tests were performed to compare the reported parameters between training groups. No statistical differences were found for sensitivity and specificity, independent of the SoT. Accuracy was higher for expert in-person trained readers than for the e-trained readers (p=0.03) (Table 1). Individual reader results of visual assessments are provided in supplementary Table 1.

Inter- and intra-reader agreement

Inter-reader agreement (Fleiss' kappa) was 0.89 (considered excellent, 95% confidence interval [CI]: 0.82, 0.97) for expert in-person-trained readers and 0.71 (considered very good, 95% CI: 0.62, 0.81) for naïve e-trained readers (Table 2). Median (range) intra-reader agreement was 0.9 (0.79–0.90) for expert in-person trained readers and 0.9 (0.66–1.00) for e-trained readers (Table 3).

Comparison of visual and quantitative assessments

81 of 82 images were evaluated with both quantitative and visual methods by the in-person trained readers and e-trained readers. Receiver operating characteristic curve analysis of composite grey matter SUVRs resulted in an optimal cut-off of 1.47 with a sensitivity of 85.7% (95% CI, 73.8–93.6%) and a specificity of 92.0% (95% CI, 74.0–99.0%. 87.7% (n=71) of the scans were classified as positive or negative by both the e-trained readers and quantitative approaches, with the remaining 12.3% (n=10) of scans showing a discrepancy between the two approaches. Both methods (visual assessment by five e-trained readers and SUVR quantification)

were significantly and strongly related (X^2=44.19, p<0.0001, Kappa = 0.73; Table 4, Figure 2A). Similarly, for in-person trained readers; 88.9% (n=72) of assessments were concordant and 11.1% (n=9) were discordant with quantitation, with significant relationship between methods (X^2 = 47.33, p<0.0001, Kappa = 0.76; Table 5, Figure 2B).

Further details of the 10 discordant cases are summarized in Table 6. Concordance with the pathology results was found with visual assessment in 9 cases by in-person trained readers, in 8 cases by e-trained readers, but with the quantitative assessment only 2 cases matched the pathology results. Marked atrophy was found in 7 cases, all with visual assessment matching pathology results (6 positive and one negative case). However, for the 7 marked atrophy cases none of the quantitative assessment results matched with pathology, showing SUVRs < 1.47 in the 6 positive cases and SUVR of 1.52 in the negative case with marked atrophy. Reader agreement for 8 of 10 cases was very high both for the in-person and electronically trained reader groups. Only for one subject (case # 75 in table 6, supplementary table 1) the visual assessments did differ for the two methods.

Challenging cases

The patient cohort included challenging cases, such as subjects with marked brain atrophy or scans with head motion. Figure 3 shows sample florbetaben PET scans from a patient with marked brain atrophy, with potential for false-positive assessment. A small percentage of scans demonstrated motion as illustrated in Figure 4.

DISCUSSION

Results from this study demonstrate high sensitivity and specificity of florbetaben PET imaging for evaluation of β-amyloid plaques in end-of-life individuals. Importantly, the sensitivity and specificity were independent of the read training method, the previous expertise

of readers in β-amyloid PET scan assessment, and the histopathology method used (BSS with or without immunohistochemistry). However, specificity increased for BSS+immunohistochemistry compared with BSS alone, as four brains in the BSS+immunohistochemistry group were additionally categorized as β-amyloid positive. Indeed, BSS+immunohistochemistry is recommended in current neuropathological guidelines for assessment of AD pathology (*10,11*). Intra-reader and inter-reader agreement was very high for both training groups.

The individual differences in sensitivity and specificity amongst the five inexperienced readers are particularly informative. There was a wide range of specificity in particular, with two readers showing relatively lower performance than the other three. Post-study interviews with each individual reader suggested that those with poorer performance were less rigid in the application of the training rules to their visual assessments. Specifically, when assessing whether a particular region is positive, the algorithm requires that the majority (i.e. at least 50%) of the axial slices comprising each region must be positive. In some instances, the poorer readers did not apply this rule systematically, leading to a region being classified as positive. Nonetheless, the sensitivity and specificity in the present study were consistent with earlier analyses from the same study (*3*) and also with previous reports using florbetaben PET (*6,12*) using clinical diagnosis as the SoT. Pathology examinations have shown, however, that clinical diagnosis of AD can be wrong (i.e. individuals diagnosed with AD dementia who do not show β-amyloid plaques upon autopsy) in 10–30 % of cases (*13*). This can lead to false-negatives compromising the sensitivity estimation when using clinical diagnosis as SoT.

The relationship of visual reads to quantitative values was also consistent with prior reports (*6,14*), indicating high concordance between visual positivity and negativity and the composite SUVR using a quantitative cut-off for positive and negative scans. There were only

few cases (10 for e-trained and 9 for in-person trained out of 81 case evaluations) in which there was discordance between visual and quantitative assessments. The visual assessment method used did not allow comparison of the PET scan images with CT or MR. This may lead to difficulties in the interpretation of some cases with atrophy, and quantification could potentially help here. Most of the visual vs quantitative discordant cases showed marked atrophy. However, in this study the visual assessment of discordant cases matched with pathology results. Quantitative assessment of cases with marked atrophy were all discordant with pathology results. The readers were more adept at distinguishing gray matter from white matter uptake in scans with severe atrophic changes than the quantification method. Two exceptions to this are shown in figures 1 and 2 in the supplementary material. This suggests that atrophy may affect the quantification method used in this study more than the visual assessment. Further investigation is required to substantiate this and whether partial volume error correction influences this result.

The end-of-life population used in this study is not the intended population for β-amyloid PET scanning. The clinically intended population will likely have less structural brain abnormalities observed in this study cohort. Furthermore, the reading methodology designed for florbetaben in the present study was quite restrictive, with readers not permitted to use all the tools routinely available in nuclear medicine for PET assessment (*3*). In clinical practice, PET scans are read using three spatial orientations, with structural images provided by CT or MRI to guide anatomical localization of findings, often in discontinuous color scales for fused images. Additional use of all available tools for image reading will likely increase the diagnostic performance. Nevertheless, excellent results were obtained even allowing for the challenging nature of some scans and the stringent requirements of the applied reading methodology.

CONCLUSION

Overall, visual assessment of florbetaben images was robust even in challenging scans from elderly end-of-life individuals. Sensitivity and specificity were high, as was inter-reader agreement, independent of the reader expertise and training method employed. The visual assessment strategy and respective training tools to analyze florbetaben PET scans are accurate and reliable in the detection of brain neuritic β-amyloid plaques as assessed using BSS, and cored plaques as assessed using immunohistochemistry, and showed a good correlation with quantitative assessments.

DISCLOSURE

John Seibyl holds equity interest in Molecular Neuroimaging and is a consultant to Piramal Imaging and GE Healthcare. Ana M Catafau, Santi Bullich, Andrew W Stephens, Andre Mueller, and Norman Koglin are employees of Piramal Imaging GmbH, Berlin, Germany. Henryk Barthel received consultant and speaker honoraria as well as travel expenses from Bayer Healthcare/Piramal Imaging. Kenji Ishii has been a paid consultant to GE Healthcare. Christopher C. Rowe has received research grants from Bayer Schering Pharma. James B. Leverenz is a consultant with Bayer, Citibank, Piramal Imaging, and Navidea Biopharmaceuticals. Bernardino Ghetti is a consultant with Piramal Imaging and has a contract with Eli Lilly and Company. James W. Ironside is a consultant with Piramal Imaging, Covance, UK, and has received honoraria from Springer and McCann Healthcare. Work by Hiroyasu Akatsu was partially supported by funding from Bayer. Masaki Takao received a research grant from Bayer Health Care. Walter J. Schulz-Schaeffer received research grants from Bayer HealthCare and Piramal Imaging. Marwan W. Sabbagh has contracts or grants with Bayer HealthCare, Piramal Imaging, Navidea Biopharmaceuticals, Avid, GE Healthcare, Avanir, Elan,

Functional Neuromodulation, Eisai, Pfizer, and Genentech, is a consultant for Lilly, Avid, Piramal Imaging, Biogen and Eisai, and receives royalties from Ten Speed and Wiley. Osama Sabri received consultant and speaker honoraria as well as travel expenses from Bayer Healthcare/Piramal Imaging. Anja Hoffmann is employee of Bayer Pharma AG, Berlin, Germany. The other authors have no disclosures.

ACKNOWLEDGEMENTS

The trial was funded by Bayer Pharma AG, Berlin (Germany), and Piramal Imaging S.A., Matran (Switzerland). Medical writing support was provided by Dan Booth (Bioscript Medical, London, UK) and funded by Piramal Imaging S.A.

REFERENCES

1. Piramal Imaging. *NEURACEQ (florbetaben F 18 injection), for intravenous use. Prescribing information*. Matran: Piramal Imaging, 2014.

2. Piramal Imaging Limited. *Neuraceq. Summary of product characteristics*. Cambridge: Piramal Imaging Limited, 2014.

3. Sabri O, Sabbagh MN, Seibyl J, *et al*. Florbetaben PET imaging to detect amyloid plaques in Alzheimer disease: phase 3 study. *Alzheimers Dement.* 2015;11:964-974.

4. Boellaard R. Standards for PET image acquisition and quantitative data analysis. *J Nucl Med.* 2009;50 (Suppl 1):11S-20S.

5. Becker GA, Ichise M, Barthel H, *et al*. PET quantification of 18F-florbetaben binding to beta-amyloid deposits in human brains. *J Nucl Med.* 2013;54:723-731.

6. Barthel H, Gertz HJ, Dresel S, *et al*. Cerebral amyloid-beta PET with florbetaben (18F) in patients with Alzheimer's disease and healthy controls: a multicentre phase 2 diagnostic study. *Lancet Neurol.* 2011;10:424-435.

7. Barthel H, Luthardt J, Becker G, *et al*. Individualized quantification of brain beta-amyloid burden: results of a proof of mechanism phase 0 florbetaben PET trial in patients with Alzheimer's disease and healthy controls. *Eur J Nucl Med Mol Imaging.* 2011;38:1702-1714.

8. Mirra SS, Heyman A, McKeel D, *et al*. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology.* 1991;41:479-486.

9. Uchihara T. Silver diagnosis in neuropathology: principles, practice and revised interpretation. *Acta Neuropathol.* 2007;113:483-499.

10. Hyman BT, Phelps CH, Beach TG, *et al*. National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimers Dement.* 2012;8:1-13.

11. Montine TJ, Phelps CH, Beach TG, *et al*. National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease: a practical approach. *Acta Neuropathol.* 2012;123:1-11.

12. Tiepolt S, Barthel H, Butzke D, *et al*. Influence of scan duration on the accuracy of beta-amyloid PET with florbetaben in patients with Alzheimer's disease and healthy volunteers. *Eur J Nucl Med Mol Imaging.* 2013;40:238-244.

13. Beach TG, Monsell SE, Phillips LE, Kukull W. Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010. *J Neuropathol Exp Neurol.* 2012;71:266-273.

14. Jennings D, Seibyl J, Sabbagh M*, et al*. Age-dependence of brain b-amyloid load in Down syndrome: a [18F]florbetaben PET study. *Neurology*. 2015;84:500-507.

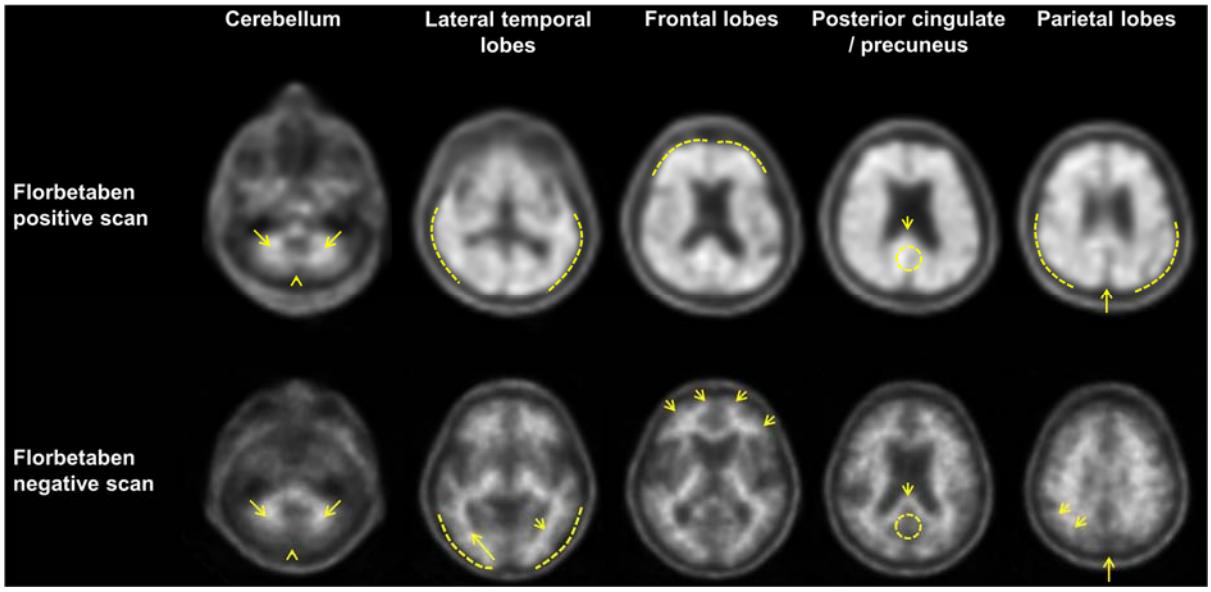

Figure 1. Visual patterns of abnormality and normality taught to readers using an electronic training tool or during in-person training. **Cerebellum:** A contrast between the white matter (arrows) and grey matter is seen in both negative and positive scans. Extracerebral tracer uptake in scalp and in the posterior sagittal sinus (arrowhead) can be seen. **Lateral temporal lobes:** The positive scan shows a "plumped", smooth appearance of the outer border of the brain (dashed line) from tracer uptake in the grey matter. Spiculated or "mountainous" appearance of the white matter (arrows) characterizes the negative scan. **Frontal lobes:** The positive scan shows the tracer uptake has a "plumped", smooth appearance due to the grey matter signal (dashed line). Spiculated appearance of the white matter in the frontal lobes (arrows) is seen in the negative scan. **Posterior cingulate / precuneus:** Adjacent to the splenium (arrow), the region appear as a hypo-intense "hole" (circle) in the negative scan, whereas this hole is absent (circle) in the positive scan. **Parietal lobes:** In the positive scan, the midline between the parietal lobes is thinner. The cortical areas are "filled-up" and show smoother appearance as the uptake extends to the outer rim. In the negative scan, the midline between the parietal lobes can be easily identified (long arrow); white matter has spiculated appearance (short arrow) with less uptake to the outer rim (dashed line).

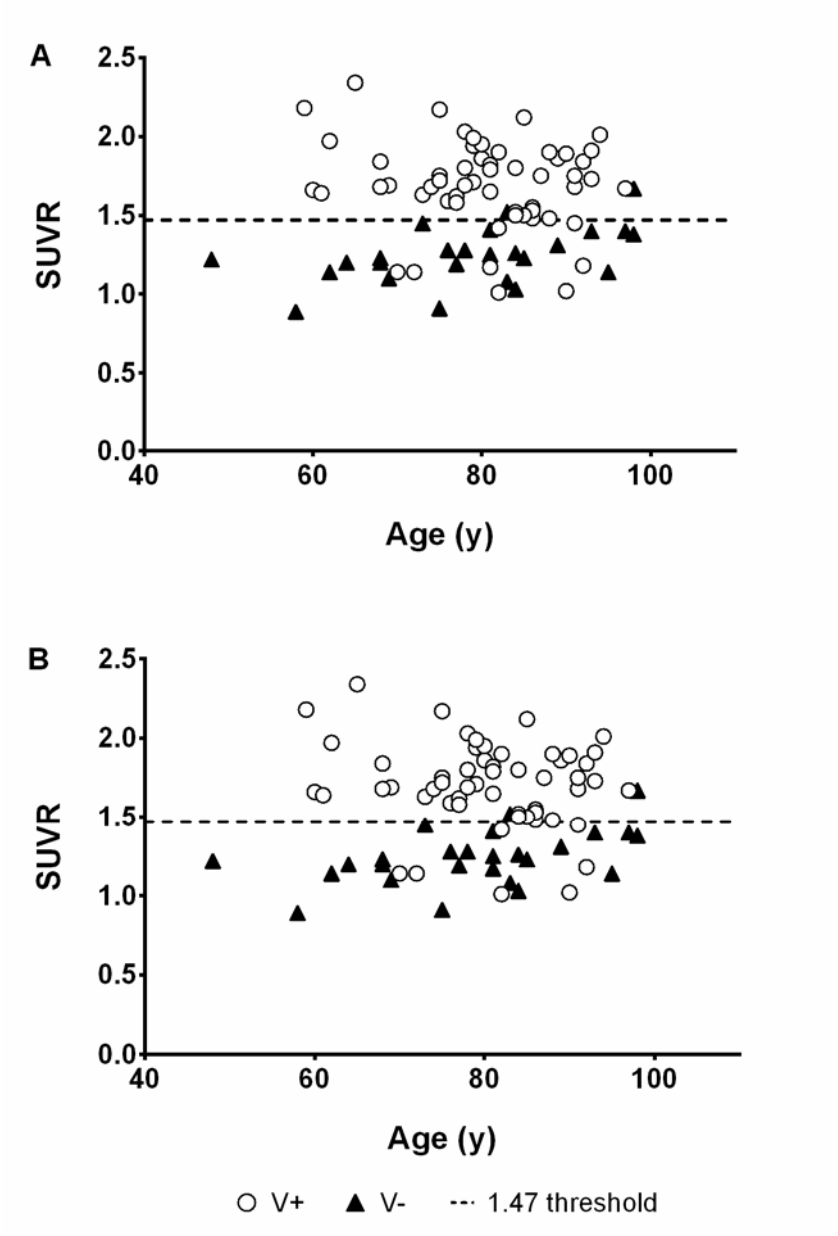

Figure 2. Comparison of visual and quantitative image assessment. Visually assessed images were based on the majority results classified into V+ (visual positive scan) and V- (visual negative scan). Assessment data is plotted related to composite SUVR and subject age (years). A) Depicts the majority read data of the electronically trained readers and, B) shows majority read data of the in-person trained readers.

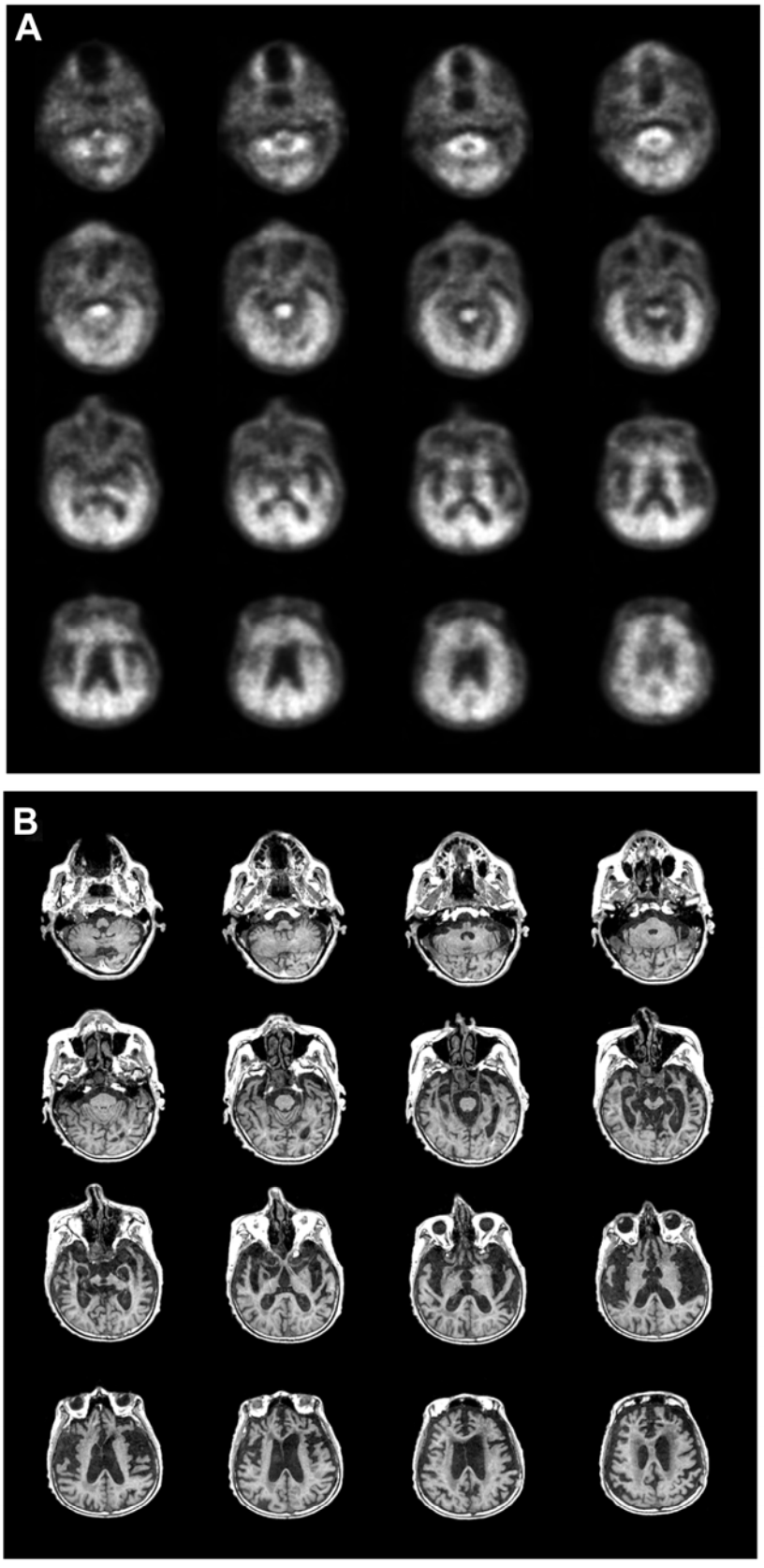
A
B

Figure 3. Florbetaben a) PET and b) MRI scans from patient # 71 with marked brain atrophy, which was assessed by all readers as β-amyloid positive. The patient was found to be β-amyloid negative on post-mortem histopathology and positive in the quantitative assessment (SUVR = 1.53).

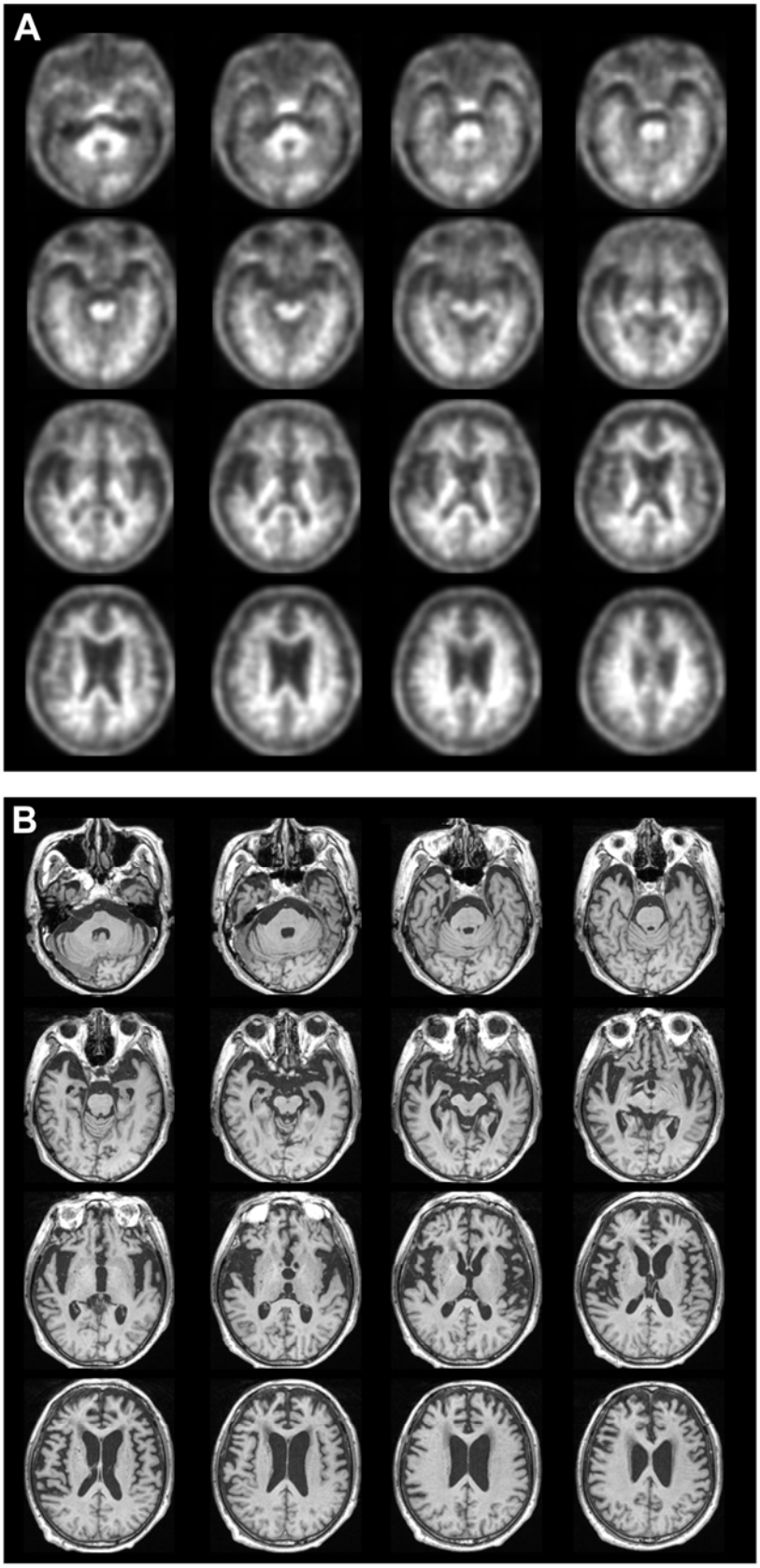
A
B

Figure 4. Florbetaben a) PET and b) MRI scans from patient # 68 with motion artifact, potentially leading to a false-positive visual assessment. All readers assessed the scans as β-amyloid negative. The patient was found to be β-amyloid negative on post-mortem histopathology and negative on the quantitative assessment (SUVR = 1.10).

Table 1. Sensitivity, specificity and accuracy of visual florbetaben PET scan reads in the eight readers (n=82 scans).

		BSS			BSS /IHC		
	Reader	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Accuracy (95% CI)**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Accuracy (95% CI)**
In-person training	**1**	96.2 (90.9-100.0)	83.3 (70.0-96.7)	91.5 (85.4-97.5)	94.6 (88.7-100.0)	92.3 (82.1-100.0)	93.9 (88.7-99.1)
	2	98.1 (94.3-100.0)	76.7 (61.5-91.8)	90.2 (83.8-96.7)	98.2 (94.7-100.0)	88.5 (76.2-100.0)	95.1 (90.5-99.8)
	3	98.1 (94.3-100.0)	80.0 (65.7-94.3)	91.5 (85.4-97.5)	98.2 (94.7-100.0)	92.3 (82.1-100.0)	96.3 (92.3-100.0)
Electronic media training	**1**	94.2 (87.9-100.0)	80.0 (65.7-94.3)	89.0 (82.3-95.8)	92.9 (86.1-99.6)	88.5 (76.2-100.0)	91.5 (85.4-97.5)
	2	98.1 (94.3-100.0)	46.7 (28.8-64.5)	79.3 (70.5-88.0)	98.2 (94.7-100.0)	53.8 (34.7-73.0)	84.1 (76.2-92.1)
	3	90.4 (82.4-98.4)	80.0 (65.7-94.3)	86.6 (79.2-94.0)	91.1 (83.6-98.5)	92.3 (82.1-100.0)	91.5 (85.4-97.5)
	4	96.2 (90.9-100.0)	76.7 (61.5-91.8)	89.0 (82.3-95.8)	96.4 (91.6-100.0)	88.5 (76.2-100.0)	93.9 (88.7-99.1)
	5	100.0	56.7- (38.9-74.4)	84.1- (76.2-92.1)	100.0	65.4 (47.1-83.7)	89.0 (82.3-95.8)
Statistics	**p-value**	0.64	0.28	0.03	0.76	0.21	0.05

BSS = Bielschowsky silver staining; IHC = immunohistochemistry; Statistics = Independent Mann-Whitney tests

Table 2. Inter-reader Cohen's kappa values* (95% confidence intervals)

	In-person training				Electronic media training					
In-person	**1**	**2**	**3**		**Electronic media**	**1**	**2**	**3**	**4**	**5**
1	x	0.93 (0.85–	0.89 (0.79–		**1**	x	0.57 (0.38–	0.87 (0.76–	0.91 (0.82–	0.63 (0.45–
2		x	0.86 (0.75–		**2**		x	0.53 (0.34–	0.66 (0.48–	0.75 (0.59–
3			x		**3**			x	0.86 (0.76–	0.58 (0.41–
					4				x	0.71 (0.56–
					5					x

*Kappa values in the range 0.81–1.00 are considered excellent, 0.61–0.80 very good, and 0.41–0.60 moderate

Table 3. Intra-reader Cohens' kappa values

	Reader	Cohen's kappa (and 95% CI)
In-person Training (n=22)	1	0.90 (0.72 - 1.00)
	2	0.90 (0.71 - 1.00)
	3	0.79 (0.51 - 1.00)
Electronic media Training (n=20)	1	1.00 (1.00 - 1.00)
	2	0.66 (0.30 - 1.00)
	3	0.90 (0.71 - 1.00)
	4	0.90 (0.70 - 1.00)
	5	1.00 (1.00 - 1.00)

CI = confidence interval

Table 4. Electronically trained visual assessment in comparison to composite SUVR quantification.

	Quantitative assessment		
	Positive	**Negative**	**Total**
Visual assessment			
Positive	48	8	56
Negative	2	23	25
Total	50	31	81

Table 5. In-person visual assessment in comparison to composite SUVR quantification.

	Quantitative assessment		
	Positive	**Negative**	**Total**
Visual assessment			
Positive	48	7	55
Negative	2	24	26
Total	50	31	81

Table 6. Comparison of visual versus (semi)quantitative analysis for discordant cases (i.e. cases visually read as positive but quantified as negative or visually read as negative and quantified as positive).

Case	Age	Visual in-person trained		Visual e-trained		Quanti-fication	SUVR	PET evaluation	Histopathology (BSS+IHC)	Clinical diagnosis
			Agreement		Agreement					
3	82	positive	100%	positive	100%	negative	1.01	Marked atrophy	positive	AD
5	90	positive	100%	positive	100%	negative	1.02	Marked atrophy	positive	AD
39	82	positive	100%	positive	100%	negative	1.42	Marked atrophy	positive	AD
40	70	positive	100%	positive	100%	negative	1.14	---	positive	Other dementia
43	92	positive	100%	positive	80%	negative	1.18	Marked atrophy	positive	AD
44	72	positive	100%	positive	100%	negative	1.14	Marked atrophy	positive	AD
45	91	positive	100%	positive	100%	negative	1.45	Marked atrophy	positive	AD
75	81	negative	67%	positive	80%	negative	1.17	---	negative	AD
22	98	negative	67%	negative	80%	positive	1.67	---	positive	AD
67	83	negative	100%	negative	80%	positive	1.52	Marked atrophy	negative	Other dementia

AD = Alzheimer's disease, Agreement = Inter-reader agreement within training group, PET evaluation = additional PET assessment conducted after initial study.